Ein Verlagsbeihefter der Heise Zeitschriften Verlag GmbH & Co. KG



Storage

Schwerpunkt: Backup-Software

Veranstaltungen

5. – 9. Januar 2009, San Francisco
Macworld Conference & Expo
www.macworldexpo.com

24. – 26. Februar 2009, Cannes
VMworld Europe 2009
www.vmworld.com/europe

24. – 27. Februar 2009, San Francisco
7th USENIX Conference on File and Storage Technologies (FAST '09)
www.usenix.org/event/fast09/

3. – 5. März 2009, Nürnberg
Embedded World
www.embedded-world.de/de/

3. – 8. März 2009, Hannover
Cebit 2009
www.cebit.de

iX extra
Storage zum Nachschlagen:
www.heise.de/ix/extra/storage.shtml



Paradigmenwechsel

Konzepte und Verfahren für die moderne Datensicherung

Vorbei sind die Zeiten, in denen der Administrator den Datenbestand allnächtlich auf Band sicherte. Heutige Geschäftsanforderungen lassen sich oft nur mit modernen Techniken wie Snapshots, Deduplizierung und CDP erfüllen. Der Tatsache müssen auch die Anbieter von Backup-Software Rechnung tragen.

Das klassische Backup verändert sein Gesicht. Immer mehr Unternehmen gehen dazu über, Daten auf Festplatte zu sichern – nicht nur als Zwischenschritt, sondern für längerfristiges Backup und sogar für die Archivierung. Dies begünstigt die Anwendung neuer Verfahren zur Datensicherung wie Snapshots, Deduplizierung und Continuous Data Protection (CDP).

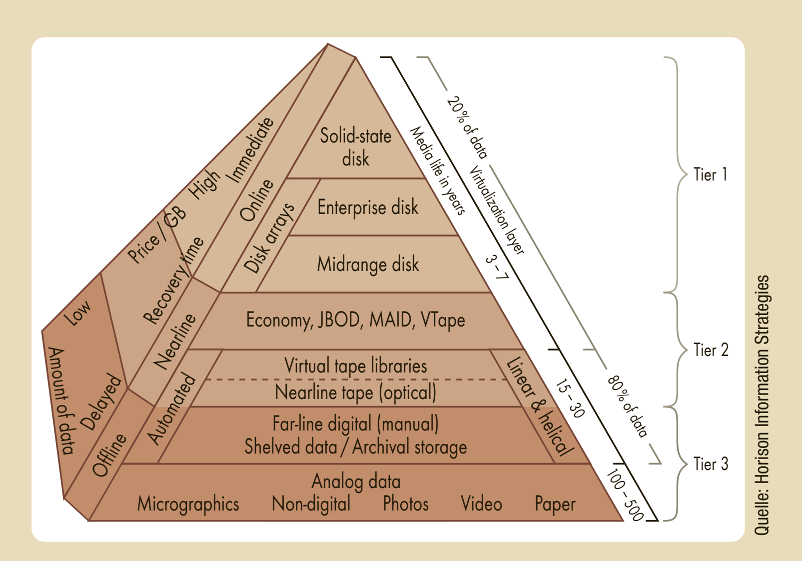
Historisch betrachtet bedeutete Backup, dass der Administrator einmal pro Woche – meist am Wochenende – eine Vollsicherung auf Tape durchführte und sie an jedem Wochentag, in der Regel nach Feierabend, durch ein inkrementelles oder differenzielles Backup der neuen und veränderten Dateien ergänzte, ebenfalls auf Tape. Viele Unternehmen praktizieren es noch heute so.

Komplexere Speicherinfrastrukturen (Tiered Storage) verändern jedoch diese Praxis. Das Band als Datenspeicher verschiebt sich in der Speicherhierarchie immer weiter nach hinten: Wo ehemals nur bei DAS (Direct Attached Storage) oder SAN (Storage Area Network) zunächst auf Platte und anschließend auf Tape gesichert wurde, hat sich heute eine Hierarchie von Online-, Nearline- und Offlinespeicher etabliert, teilweise mit mehreren Unterstufen (siehe Abbildung 1). Jedes Unternehmen hat eine individuelle Ausformung dieser Hierarchie im Einsatz, ob es ihm bewusst ist oder nicht. Rechnet man Solid State Disks (SSD) als Zusatz-Cache für besonders schnellen Datenzugriff (Tier 0) mit, kommen viele Anwender auf eine Hierarchie mit fünf oder mehr Stufen.

Disk ist schneller als Tape

Für diese Entwicklung gibt es diverse Gründe. Dazu zählen geschäftliche Anforderungen, etwa bestimmte Anwendungsdaten lange im Onlinezugriff zu halten, wobei man aufgrund der technischen Entwicklung einen kleineren Anteil auf hoch-

performanten und teuren Fibre-Channel- oder SAS-Platten belässt und einen größeren auf weniger leistungsfähige, aber billigere SATA-Platten verschiebt. Von dort sichert man die Daten häufig auf eine Virtual Tape Library (VTL), weil das Kopieren von Disk zu Disk komfortabler und schneller ist als von der Platte direkt aufs Band. Die gefürchteten (zu) kleinen Backup-Fenster lassen sich so besser nutzen. Eine wachsende Zahl von Unternehmen verzichtet mittlerweile darauf, die Backups anschließend auf eine echte Tape Library auszulagern. Selbst eine Archivierung auf Bändern findet nicht mehr in jedem Fall statt.



Storage-Pyramide: Eine Speicherhierarchie kann etliche Stufen enthalten, die sich in puncto Kosten, Performance, Kapazität und Aufbewahrungszeit unterscheiden (Abb. 1).

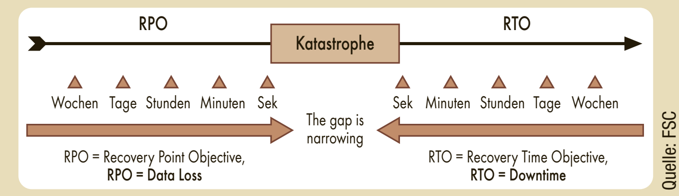
Nach Ansicht der Gartner-Analysten Dave Russell und Carolyn DiCenzo liefert Tiered Storage nicht nur eine Vielzahl von Performance- und Kostenstufen für die zu speichernden Daten, sondern stellt zunehmend auch unterschiedliche Ansätze für Backup und Recovery zur Verfügung [1]. Neue Verfahren zur Datenreduzierung und Replikation, geringere Übertragungskosten durch alternative Verbindungstechniken wie iSCSI oder Fibre Channel over Ethernet (FCoE) sowie der Aufbau größerer und kostengünstigerer Speicherumgebungen (Grids) dürften das Outsourcing von Backup und Recovery attraktiver machen, wie Amazons Simple Storage Service (S3) oder EMCs Mozy schon heute zeigen.

IT ist nur in seltenen Fällen ein Profit Center, das direkt zu Umsatz und Gewinn eines Unternehmens beiträgt. Im normalen Alltag unterstützt sie die Geschäftsabläufe und erhöht durch das Wegrationalisieren menschlicher Arbeitskraft oder die Beschleunigung von Arbeitsprozessen die Produktivität. Insgesamt verursacht die unternehmensinterne IT jedoch Kosten, die sich gerade bei Prozessen wie der Datensicherung nicht schönreden lassen.

Wenig Mittel fürs ungeliebte Kind

Praktische Folge sind Widerstände bei der Bereitstellung finanzieller Mittel für Szenarien, von denen man nicht weiß, ob sie jemals eintreten. Der Backup-Beauftragte im Unternehmen muss häufig all seine Überredungskünste nutzen, um überhaupt eine Minimalinstallation für Backup und Recovery bereitstellen zu können. Manche freut es sogar, wenn Daten verloren gehen: Nur so können sie durch eine gelungene Wiederherstellung den Nachweis erbringen, dass das in die Datensicherung investierte Geld sinnvoll angelegt ist.

Für geschäftskritische Anwendungen wie SAP oder Oracle, die eine unmittelbare Wiederherstellung des ursprünglichen Zustands verlangen, ist die traditionelle tägliche Datensicherung nicht mehr ausreichend. Wie Dave Russell im Gespräch darlegt, zeigen Kundenbefragungen, dass immer mehr Anwender den Nutzen der herkömmlichen Bandsicherung anzweifeln. Als Grund für diese Entwicklung sieht er die ständigen Änderungen, denen geschäftskritische Daten unterworfen sind.



Zeitpunkte: Anforderungen an ein Backup/Recovery-Verfahren lassen sich mit den Parametern RPO und RTO charakterisieren (Abb. 2).

Tägliche Sicherung hat ausgedient

Es gibt nicht mehr nur einen Sicherungs- oder Recovery-Punkt, sondern viele, die sich über den ganzen Tag verteilen. Unter diesen Bedingungen nur einmal täglich zu sichern, bedeutet bei einem Ausfall den Verlust der Arbeit von bis zu 24 Stunden. Immer mehr Unternehmen setzen als vorbeugende Maßnahme hostbasierte Snapshots in Intervallen von einigen Minuten oder Stunden ein oder sichern alle zwei bis drei Stunden den Datenbestand auf sekundären Plattenspeicher.

Tape und nächtliche Backups können für bestimmte Applikationen ausreichend sein, aber für Daten, die länger auf Tier 1 liegen, sollte man mehr tun, meint Russell. Es gibt heute ausgefeilte Recovery Tiers und Verfahren beziehungsweise Software, die Datenverluste begrenzen. Außerdem lassen sich Systeme schneller wiederherstellen als beim umständlichen Wiedereinspielen von Sicherungsbändern.

Aus den Geschäftsanforderungen lässt sich die sogenannte Recovery Point Objective (RPO) ableiten. Sie gibt die maximale Zeit zwischen einem Ausfall und der letzten Datensicherung an. Gleichzeitig ist sie ein Maß für die Höhe der zu erwartenden beziehungsweise tolerierbaren Datenverluste – je länger die letzte Sicherung zurückliegt, desto mehr Daten gehen verloren. Unter Recovery Time Objective (RTO) hingegen versteht man die Zeit, die man zur vollständigen Wiederherstellung nach einem Ausfall benötigt (siehe Abbildung 2).

RAID-Pannen und Katastrophen wie Feuer, Überschwemmungen oder terroristische Angriffe kommen vergleichs-



Storage - worldwide with Open-E

ab 760,00 €

ab 539,00 €

ab 499,00 €

Unverbindliche Preisempfehlung (Endkundenpreis inkl. MwSt.)

Open-E Software, ein optimiertes Betriebssystem für Storage Systeme

- Gemeinsamer Dateizugriff
- Speicherzusammenlegung
- Backup und Wiederherstellung
- Storage-Virtualisierung
- Replizierung und Archivierung
- High Available Storage für Cluster
- Speicherkapazität, Standard 4, 8, 16 TB
- einfache Erweiterbarkeit
- große Ausfallsicherheit
- hohe Verfügbarkeit
- einfache Administration
- sehr gutes Preis-Leistungs-Verhältnis

ADN
Advanced Digitel
Network Distribution GmbH
T: +49 (2327) 9912 - 0
F: +49 (2327) 9912 - 540
info@adn.de
www.adn.de



CPI
Computer Partner Handels
GmbH
T: +49 (89) 9624410
F: +49 (89) 96244135
sales@cpigmbh.de
www.cpigmbh.de



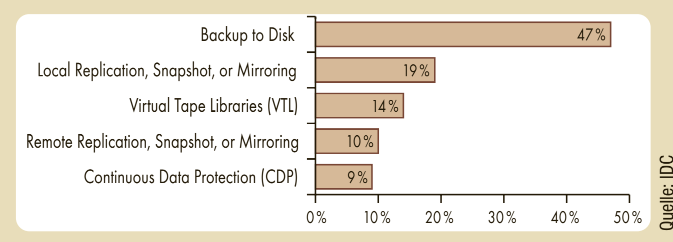
CTT AG
T: +49 (89) 4209000
F: +49 (89) 42090099
info@ctt.de
www.zero-one.de



MICROTRONICA
T:+ 49 (6103) 3048214
F:+ 49 (6103) 3048581
info@microtronica.de
www.microtronica.de



www.open-e.com www.open-e.com/partner www.open-e.com/free



Aufholbedarf: Nach einer IDC-Umfrage nutzen nur wenige europäische Unternehmen die Möglichkeiten, die ein festplattenbasiertes Backup bietet (Abb. 3).

weise selten vor. Das Gros aller Recovery-Fälle entsteht durch Anwender- oder Softwarefehler – laut Gartner etwa 80 Prozent. Ihnen kann man mit regelmäßigen Snapshots beikommen, die den aktuellen Zustand des Dateisystems „einfrieren“ und eine schnelle Wiederherstellung auch einzelner Dateien erlauben. Wählt man ausreichend kurze Snapshot-Intervalle, lässt sich eine RPO von einigen Minuten erreichen.

CDP verringert Datenverluste

Das „Einfrieren“ der Daten auf dem Host hilft allerdings nicht weiter, wenn das gesamte System ausfällt und man nach der Reparatur wieder auf den Stand der letzten, weiter zurückliegenden Sicherung ausweichen muss. Continuous Data Protection (CDP) geht einen Schritt weiter: Sie erfasst alle Änderungen und kopiert sie auf ein zweites Speichermedium. Im Gegensatz zur Replikation, die nur den aktuellen Stand der Daten vorhält, katalogisiert sie jedoch alle Änderungen, sodass man auch auf ältere Dateiversionen zugreifen kann. Da jeder Schreibvorgang umgehend ins Backup übernommen wird, ist die Recovery Point Objective praktisch Null.

Snapshots wurden zuerst in der Mainframe-Welt von IBM und Storagetek (heute Sun Storagetek) entwickelt, während bei den „Open Systems“ EMC, Netapp oder jüngst Compellent vergleichbare Verfahren entwickelten. Die Bemühungen der Softwareproduzenten gehen laut Russell dahin, zusätzliche Funktionen in die Programme hineinzuverlagern, aber erst in etwa fünf Jahren sei so etwas wie ein Unified Recovery Management zu erwarten.

Neue Techniken selten genutzt

Backup-Anbieter wie Commvault, EMC oder Symantec arbeiten daran, das Management von Arrays mit Disk-to-Disk-Techniken, Snapshots oder der Integration von NAS-Filern zu verbinden. Start-ups wie Topio oder Timespring, die CDP-Produkte entwickelt hatten, wurden von Netapp beziehungsweise Doubletake übernommen. Während der Einsatz von Snapshots bereits recht verbreitet ist, gilt dies nicht für Continuous Data Protection. Viele Anwender geben freimütig zu, den Begriff noch nie gehört zu haben.

Redundanz im Überfluss

Backup ist die Disziplin mit der höchsten Redundanz und der größten Wiederholrate, die man heute in Unternehmen findet: Alles, was sich an Daten auf den Online-Speichersystemen befindet, wird noch einmal in eine neue Instanz übernommen – einschließlich aller vorhandenen Kopien. Das nächste Full Backup sichert alle Daten ein weiteres Mal, obwohl sich in der Regel über 90 Prozent der Dateien im Laufe ihres Lebens gar nicht mehr ändern. Der selbstständige Analyst Josh Krischer schätzt, dass heute von jeder in einem Unternehmen erzeugten Datei durchschnittlich sieben Kopien existieren, die das Backup immer wieder sichert. Deduplizierung kann die unnötige Redundanz der gesicherten Daten verringern – vorausgesetzt, man sichert auf Festplattensysteme. Bislang machen allerdings nur wenige Anwender Gebrauch davon.

Deduplizierung, seit Ende 2006 von einigen Herstellern propagiert und durch diverse Firmenübernahmen unterfüttert – Avamar ging an EMC, Rocksoft an ADIC beziehungsweise Quantum, Diligent an IBM –, ist eine Weiterentwicklung von Single Instance Storage (SIS). Das Verfahren verringert den Speicherplatzbedarf, indem es Dateien lediglich einmal speichert und alle Kopien durch Verweise ersetzt. SIS kommt unter anderem bei Mailsystemen wie Exchange, Archivprodukten wie EMCs Centera oder bei Lösungen für den Datenschutz wie Connected Backup von Iron Mountain zum Einsatz.

ANBIETER VON BACKUP- UND RECOVERY-PRODUKTEN

Die folgende Übersicht erhebt keinen Anspruch auf Vollständigkeit.

Hersteller	Produkt	Website
Acronis	Acronis True Image	www.acronis.com
Archiware	PresSTORE Backup	www.archiware.de
Atempo	Time Navigator	www.atempo.com
Bakbone	NetVault	www.bakbone.com
CA	ARCserve Backup, CA Recovery Management	www.ca.com
Commvault	Simpana	www.commvault.com
Data Domain	OpenStorage, Replicator	www.datadomain.com
EMC	NetWorker	www.emc.com
Falconstor	Continuous Data Protector, Single Instance Repository	www.falconstor.com
FSC	NetWorker Fast Start	www.fujitsu-siemens.com
HDS	ProtecTIER	www.hds.com
HP	Data Protector	www.hp.com
IBM	Tivoli Storage Manager (TSM)	www.ibm.com
Netapp	Ontap, Snap Manager	www.netapp.com
Quantum	StorNext	www.quantum.com
Symantec	Backup Exec, NetBackup	www.symantec.com
Syncsort	Backup Express	www.syncsort.de

Statt mit ganzen Dateien arbeitet Deduplizierung mit Bruchstücken davon, sogenannten „Chunks“. Die Verfahren der einzelnen Anbieter arbeiten mit Chunks unterschiedlicher – manchmal variabler – Größe. Außerdem verwenden sie unterschiedliche Algorithmen für das Zerlegen der Daten, das Suchen nach identischen Chunks und das Reduzieren der Datenmenge [2]. Viele Hersteller geben eine erreichbare Platzersparnis zwischen 20:1 und 400:1 an. Sie hängt allerdings stark von der eingesetzten Backup-Methode ab. Zum Beispiel wird die Einsparung umso höher ausfallen, je öfter man eine Komplettsicherung durchführt.

Außerdem kann ein Disk-Backup mit Deduplizierung die für die Wiederherstellung benötigte Zeit verkürzen. Je länger die gesicherten Daten auf der Festplatte bleiben können, desto geringer ist die Wahrscheinlichkeit, dass man für ein Restore auf die erheblich langsameren Sicherungsbänder zurückgreifen muss.

Modernes Backup erfordert Festplatten

Durch den Einsatz von Backup to Disk und Deduplizierung kann eine IT-Abteilung die Beeinträchtigung des täglichen Betriebs durch die Datensicherung reduzieren. Doch viele Unternehmen sind noch nicht den entscheidenden Schritt gegangen, Festplatten für Nearline- und Backup-Umgebungen einzusetzen. Selbst die, die bereits Disks verwenden, stellen oft nur wenig Plattenkapazität für ihre Datensicherung bereit oder nutzen Festplatten ausschließlich als Zwischenstation (Staging Disk), von der sie die Daten im Anschluss an das eigentliche Backup auf Band überspielen. Das zeigt eine von IDC im Jahr 2007 unter 600 europäischen Unternehmen durchgeführte Umfrage (siehe Abbildung 3).

Es bleibt abzuwarten, wie viele Anbieter dem Druck der Anwender, neue Funktionen in ihre Software zu integrieren, standhalten werden. Einige könnten noch mehr zu Nischenspielern werden – zumindest aus der Sicht von Gartner. Das Unternehmen hat in einer aktuellen Untersuchung [3] neun Anbieter – Atempo, Bakbone Software, CA, Commvault, EMC, HP, IBM, Symantec und Syncsort – unter die Lupe genommen und bewertet. Im Gegensatz zu früheren Betrachtungen haben die Analysten neue Kriterien angewendet, darunter die Unterstützung von plattenbasierter Wiederherstellung, Virtual Tape und Deduplizierung, Support für Umgebungen mit Server-Virtualisierung sowie die Einbindung von Zweigstellen ins zentrale Backup-System.

Komplettlösungen sind gefordert

Einige Hersteller neigen laut Gartner noch zu isolierten Angeboten. Der Umbau von Backup-Angeboten zu kompletten Recovery Management Tools, die sich weniger um Magnetbänder und einfache Disk-Lösungen als um anwendergerechte Software mit RPO- und RTO-Vorgaben kümmern, stehe erst am Beginn. Backup-Software für Unternehmen, die sich diesen neuen Anforderungen nicht stelle, werde im Wettbewerb zurückbleiben und in einer Nische verharren. Außerdem haben inzwischen weitere Hersteller das Terrain betreten. Zu den Angeboten zählen netzgestützte Replikation (Brocade, F5 Networks, Riverbed), eigene Snapshot-Funktionen (Netapp) oder dedizierte Deduplizierungsprodukte wie die von Data Domain. Hinzu kommen Anbieter von Software as a Service (SaaS) und Storage as a Service.

Anwender müssen die eigenen Anforderungen klar formulieren, damit sie die vorhandenen Angebote bewerten können. Neben dem Funktionsumfang und Parametern wie RPO/RTO sollten sie dabei auch „weiche“ Faktoren wie Wartung, Support und Lizenzgestaltung nicht außer Acht lassen. (mr)

Hartmut Wiehr
ist Fachjournalist in München und Herausgeber des Storage Compendium – Das Jahrbuch 2006/2007.

Literatur

[1] Dave Russell, Carolyn DiCenzo; Key Issues in Data Protection, 2008; Gartner Research, März 2008
[2] Nils Haustein; Massenspeicher; Weniger ist mehr; Wie Deduplizierung funktioniert; *iX* 3/2008, S. 135
[3] Dave Russell, Carolyn DiCenzo; MarketScope for Enterprise Backup/Recovery Software, 2008; Gartner Research, April 2008

Storage „Made in Germany“

rapidNAS

rapidNAS® DSS Lösungen zeichnen sich durch die besonders einfache und schnelle Inbetriebnahme aus.
Durch die intuitive grafische Oberfläche und eine ausgezeichnete Skalierbarkeit heben sich die rapidNAS® DSS Systeme von anderen Lösungen deutlich ab.

- Univeral Data Storage - iSCSI und NAS in einem Gerät
- Automatischer Failover
- SNMP
- Web Management GUI
- Share Synchronisation und Volume Replication
- bis 24.0 TByte Kapazität brutto
- SMTP
- unterstützt SAS und SATA HDD (auch gemischt)

rapidNAS DSS 104, 1 HE, 4 Bay
- max. 4.0 TByte brutto
- Single Netzteil, 400 Watt

rapidNAS DSS 208, 2 HE, 8 Bay
- max. 4.0 TByte brutto
- red. Netzteil 2x 500 Watt

rapidNAS DSS 316, 3 HE, 16 Bay
- max. 4.0 TByte brutto
- red. Netzteil 3x 350 Watt



rapidNAS DSS 524, 5 HE, 24 Bay
- max. 24.0 TByte brutto
- red. Netzteil 3x 500 Watt





Kosteneffizienter Speicher für Datenbankumgebungen und Servervirtualisierung

Durch die Möglichkeit SAS und/ oder SATA Festplatten verwenden zu können, findet die ICEBOX®-SAS Serie in den unterschiedlichsten Bereichen ihre Anwendung.

Ob als Speichererweiterung mit SAS Festplatten für Datenbankanwendungen und zentrales Storage für Servervirtualisierung oder zur schnellen Speichererweiterung mit günstigen SATA Festplatten für File- und Backup Server.

- 2x FC 4GB oder 4x 1GB iSCSI oder 2x SAS Interface
- Snapshot-on-the-Box
- SNMP
- Web Management GUI
- 2x 500 Watt red Netzteil
- Erweiterbar um bis zu 3 JBOD 16-bay Einheiten
- bis 64 TByte Kapazität brutto
- SMTP
- unterstützt SAS und SATA HDD (auch gemischt)
- Securion Interface (Einbindung in Alarmsysteme)

ICEBOX-FQS 16 Bay
- max. 16.0 TByte brutto
- red. Netzteil 2x 500 Watt



ICEBOX-ISQS 8Bay
- max. 8.0 TByte brutto
- red. Netzteil 2x 500 Watt





N-TEC GmbH • Oskar-Messter-Str. 14 • D-85737 Ismaning
www.ntecgmbh.de • info@ntecgmbh.de
Phone +49 (0)89 9584070 • Fax +49 (0)89 95840711

Nur keine Hierarchien

Backup im Unternehmen – ein ungeschminktes Beispiel

Unternehmen, die über genügend Expertise und Kaufkraft verfügen, halten Abstand zu den Marketingübungen der Hersteller. Da sie wissen, was sie benötigen, sind sie in der Lage, Dienstleistungen statt Versprechen einzukaufen. Das zeigt ein Beispiel aus der Praxis.

Beim Autozulieferer ZAG aus dem Ruhrgebiet, dessen wahrer Name auf eigenen Wunsch ungenannt bleiben soll, versteht sich das betriebseigene Rechenzentrum als Dienstleister, der Fachabteilungen und Standorte in Deutschland und im europäischen Ausland zu seinen „Kunden“ zählt. Innerhalb des zentralen Rechenzentrums ist eine Arbeitsgruppe für die Backup-Konzepte sowie für Kontakte und Verhandlungen mit den Herstellern zuständig, während die Umsetzung im Verantwortungsbereich der Applikationsbetreuer liegt.

Innerhalb der IT-Infrastruktur finden sich bei dem Unternehmen noch einige Mainframes. Doch ist ihre Bedeutung inzwischen geringer als die der „Open Systems“: der Unix- und Windows-Server. Die Storage-Infrastruktur ist weitgehend in einem SAN konsolidiert, wobei ZAG Produkte zweier Hersteller einsetzt: EMC und Hitachi Data Systems (HDS). Das zahlt sich bei Verhandlungen mit den beiden Lieferanten aus, wenn es um bestimmte inhaltliche Forderungen bei den Produkten geht, aber auch bei Lizenzen und Preisen hält man einen Hebel in der Hand. Praktische Folge: Beide Hersteller sind etwa zu gleichen Teilen im Unternehmen vertreten. Zusätzlich setzt ZAG Systeme von Netapp als Fileserver ein. Für die Datensicherung kommt flächendeckend der Tivoli Storage Manager (TSM) von IBM zum Einsatz.

Klassenkampf der Dateien

Die IT-Abteilung garantiert für bestimmte Services definierte Verfügbarkeiten und Anlaufzeiten. Dies schlägt sich bei Backup und Restore in drei unterschiedlichen Storage- und Dateiklassen nieder, die genau festlegen, wie weit man etwa im Backup zurückgehen kann. Getrennt für Datenbanken und andere strukturierte Daten beziehungsweise unstrukturierte Dateien hat das Unternehmen drei Verfügbarkeitsstufen aufgebaut:

– 1. Hochverfügbarkeit auf Systemen mit hoher Performance, gespiegelt, mit langen Aufbewahrungszeiten beim Backup;
– 2. fast gleich hohe Performance, jedoch ohne Spiegelung, Backup mit reduzierten Aufbewahrungszeiten;
– 3. geringere Performance, ungespiegelt, stärker reduzierte Aufbewahrungszeiten für Backups.

Für beide Datenklassen hat ZAG eine gemeinsame SAN-Umgebung eingerichtet. Dabei sind die Fileserver aufgrund ihrer „autistischen“ Natur teilweise separat angesiedelt. „Sie kennen nur sich selbst, schauen weder rechts noch links“, beschreibt der verantwortliche IT-Leiter die Schwierigkeiten, sie in eine gemeinsame Speicher- und Backup-Infrastruktur zu integrieren. Das sei durch die Konstruktion der NAS-Filer bedingt. Man wünsche sich von Netapp, dass der Anbieter ein globales Dateisystem einführt, das die Integration in die Backup-Umgebung erleichtert.

Netapp und Tivoli/IBM reagieren: Im nächsten Jahr wollen die Unternehmen eine gemeinsame Schnittstelle mit der Bezeichnung „Snapdiff“ herausbringen. Darauf, auf eine Standardisierung durch die Storage Networking Industry Association (SNIA) zu warten, habe ZAG zugunsten einer direkten Einflussnahme auf die Hersteller verzichtet. Dazu geht man bei dem Unternehmen auch Allianzen mit anderen Anbietern oder Partnern aus der Automobilbranche ein – in der Gewissheit, dass eine gebündelte Kaufkraft allemal ein gutes Argument ist.

HSM lohnt sich nicht immer

Eine hierarchische Speicherstruktur mit mehreren Festplattensystemen „hintereinander“ hat die IT-Abteilung abgelehnt, weil die dem Kunden garantierte Zuverlässigkeit auch für die nachgeordneten Systeme gewährleistet sein müsste. Sind die Daten auf günstigere Platten verschoben, müsste man sie nach diesem Ansatz noch einmal spiegeln. Das würde zusätzliche Kosten verursachen.

Es kommt nach Ansicht der IT-Fachleute bei ZAG hinzu, dass fast alle hierarchischen Speichersysteme auf Windows-Dateisystemen beruhen. In dieser Umgebung könne man letztlich nur 15 bis 20 Prozent der Daten wirklich migrieren. Obwohl man weiß, dass eigentlich 80 Prozent der Daten inaktiv sind, lassen sie sich nicht automatisch verlagern: Die Informationen über den letzten Gebrauch (last use) einer Datei sind nicht zuverlässig. Die migrierten 15 bis 20 Prozent noch einmal zu spiegeln, wäre ineffektiv und obendrein zu teuer.

Redundanz ist kostspielig

Es sei einfacher, die Daten auf einem Primärsystem zu halten und dem Kunden die Wahl zwischen unterschiedlichen Datentöpfen zu geben – mit hoher und geringerer Verfügbarkeit sowie mit einer Verfügbarkeit, die sogar Datenverluste in Kauf nimmt. Die Kosten für das Vorhalten eines Großteils der Daten auf einem teuren gespiegelten Primärspeicher nimmt man hin. Der zuständige Mitarbeiter kontert auf Nachfrage: „Es ist anders nicht billiger.“ Beim Primärspeicher müsse man nicht alles spiegeln, um eine hohe Verfügbarkeit zu gewährleisten. Man könne auch Datenklassen definieren, die zwar dort liegen, aber nicht gespiegelt werden. Letztlich sei das eine Frage des vom Kunden gewünschten Preis-Leistungs-Verhältnisses.

Im Backup-Umfeld hat man noch wenig Tiered Storage installiert. Normalerweise schiebt TSM die Daten auf einen Disk-Cache und migriert sie tagsüber unter Benutzung von Disk Libraries auf Band. Das sei bei dem gewählten Konzept die einzige Möglichkeit, ein Tiered Storage einzurichten, da man die Daten zunächst auf einem sehr performanten Speichersystem ohne Mount-Zeiten hat.





Setzen Sie auf CPI Server-Systeme mit energieeffizienten Komponenten. Sichern Sie sich schon heute einen hohen ROI durch Stromeinsparung von bis zu 100 Watt/h pro System im Vergleich zu einem derzeit aktuellen Standard-Server-System.

Dies erreichen wir für Sie durch den Einsatz von hochwertigen Komponenten:

· 2.5" S-ATA oder SAS Festplatten
· hocheffizienten Netzteilen mit einer Effizienz von bis zu 93%
· größere LowVoltage Speichermodule (z.B. 4 GB statt 2 GB)
· LowVoltage Prozessoren von Intel® oder AMD

Damit läßt sich eine dauerhafte Einsparung für jeden Anwender im Entry- bis Enterprisesegment realisieren.

Gerne konfigurieren wir Ihren persönlichen Server mit energieeffizienten Komponenten.



CPI Computer Partner Handels GmbH
Kapellenstr. 11
D - 85622 Feldkirchen/München

Telefon +49 (0)89 - 96 24 41 - 0
Telefax +49 (0)89 - 96 24 41 - 33
Hotline 0800 - 100 82 69
E-mail sales@cpigmbh.de



www.cpigmbh.de

Intern hat sich beim Backup im Unternehmen in den letzten Jahren nicht viel geändert, da man schon seit Jahren mit TSM auf dieselbe Software setzt. Es besteht ein grundsätzlicher Anspruch an die Leistungsfähigkeit und die Performance der Systeme, der mit dem regelmäßig erweiterten TSM erfüllbar sei.

Starkes Wachstum des Datenvolumens

Überdies gelte, dass mit Storage-Systemen, die schon vier Jahre alt sind, heute kaum mehr Staat zu machen sei, berichtet der IT-Leiter. Dies liege vor allem am extremen Datenwachstum, das zu bewältigen sei und das an die physischen Grenzen der Altsysteme gehe. Man bekommt die Daten beim Backup zum Teil nur noch schnell genug durch die Leitungen, wenn man massiv parallel arbeitet. Jede Nacht hat das ZAG-Rechenzentrum allein in Deutschland 70 bis 80 TByte an inkrementellen Daten zu sichern.

Backup ist ein „notwendiges Übel“, das man aber, so der IT-Leiter, professionell und ohne eine Laissez-faire-Haltung betreiben müsse, selbst wenn es einiges koste. Man müsse eben wie bei einer Lebensversicherung relativ viel Geld bezahlen, obwohl man sie eigentlich gar nicht in Anspruch nehmen möchte.

Kleine und große Katastrophen

Apropos Lebensversicherung: Ein teilweises Restore kommt häufig vor, weil irgendein Mitarbeiter eine Datei verloren hat oder sie nicht mehr findet. Die SAP-Administratoren haben einen größeren Restore-Bedarf – nicht weil sie Daten verlieren, sondern weil sie ihre Test- oder Qualitätssysteme meistens aus einer Wiederherstellung aufbauen. Der Vorteil für die Backup-Crew besteht darin, dass sie dies zugleich für eine Restore-Übung nutzen können.

Komplette Systemwiederherstellungen kommen maximal fünf bis sechs mal pro Jahr vor. Der Anlass ist gewöhnlich, dass ein ganzes Dateisystem verloren gegangen ist. Man müsse diese Handvoll Ausfälle allerdings zu einer sehr großen Umgebung von etwa 4000 Servern in Beziehung setzen. „Insgesamt passiert sehr wenig“, meint der zuständige Sachbearbeiter. Immerhin könne man in diesen wenigen Fällen beweisen, dass man durchaus die verlorenen Daten wiederherstellen kann.

Außerdem, versichert der Backup-Spezialist, veranstalte man natürlich jedes Jahr einmal gründliche Tests der Backup-Szenarien. Wie bei einer Feuerwehrübung wird überprüft, ob alles funktioniert – zum Beispiel das Wiederherstellen eines Servers und seiner Applikationen aus dem Backup. Im Ernstfall, wenn man etwa nachts um drei Uhr aus dem Bett gerissen wird, weil ein Restore notwendig ist, ist einfach keine Zeit mehr für ein langes Studium der Handbücher.

Für TSM als Backup-Lösung hat sich das Unternehmen aus mehreren Gründen entschieden. Es handele sich um eine Software, die man mit relativ wenig Leuten in einer großen Umgebung managen kann. Außerdem ist sie hochgradig automatisiert und kann ein sogenanntes „Progressive Incremental Backup“ durchführen (siehe Kasten). Diese Backup-Variante sichert jede Datei und jede Version nur einmal. Momentan hat man allein im Backup-Bereich etwa zwei Petabyte an physischem Speicherplatz belegt, was etwa vier Petabyte an logischen Daten entspricht. Verwendete man ein konventionelles Backup-Konzept mit Tages-, Wochen-, Monats- und Vollsicherungen, hätte man ein Vielfaches dieser Datenmenge zu speichern. Allein die Kosten für die nötigen Magnetbänder wären erheblich.

Technische Neuerungen

Die Frage nach einer weiteren Reduktion der Datenmengen durch Deduplizierung scheint sich in solch einem Kontext erst einmal nicht zu stellen. Die Backup-Experten bei ZAG haben hierzu eine eigene Position entwickelt: Man verfolgt eine doppelte Strategie, bei der man sich im Rechenzentrum selbst erst einmal dazu entschlossen hat, sie nicht anzuwenden, weil man Dedupli-

Progressive Incremental Backup

Eine fortlaufende Datensicherung lässt sich auf mehrere Arten durchführen. Die einfachste besteht darin, regelmäßig alle Daten zu sichern (Full Backup). Allerdings hat das Verfahren zwei gravierende Nachteile: Vollsicherungen benötigen viel Speicherplatz und dauern lange. Ersterem kann man durch den Einsatz von Deduplizierungstechniken begegnen. Der Zeitbedarf lässt sich damit jedoch nicht reduzieren. Im Gegenteil, er kann – abhängig vom verwendeten Verfahren – sogar noch steigen.

Als Ausweg bieten sich differenzielle oder inkrementelle Backups an. Beide sichern nur neue Dateien und solche, die sich seit der letzten Sicherung (inkrementell) oder dem letzten Full Backup (differenziell) geändert haben. Da die Mehrheit der Daten – oft über 90 % – sich nach dem Erstellen nicht mehr ändert, dauert ein inkrementelles Backup nur einen Bruchteil der Zeit, die man für eine Vollsicherung benötigt. Allerdings dauert ein Restore umso länger, je mehr inkrementelle Sicherungen der Administrator einspielen muss.

Differenzielle Sicherungen vermeiden diesen Nachteil. Dafür erkauft man sich einen anderen: Der Umfang des Backups wächst, wenn man sich vom Stand der letzten Vollsicherung entfernt. Ebenso die Zeitspanne, die man für die Sicherung benötigt. Gelegentlich muss der Admin daher ein Full Backup einschieben. Üblich sind Vollsicherungen am Wochenende und inkrementelle oder differenzielle Backups nach Feierabend.

Wiederherstellung unvollständig

Beim Restore von partiellen Sicherungen kann es zu Schwierigkeiten kommen. Tritt während der Wiederherstellung ein Fehler auf, lassen sich die folgenden Sicherungen in der Regel nicht mehr sauber einspielen. Als Folge können Dateien unter falschem Namen oder im falschen Verzeichnis auftauchen, oder bereits gelöschte wieder auf der Festplatte landen. Von der üblicherweise geforderten vollständigen und korrekten Wiederherstellung aller Daten kann daher keine Rede sein.

Auch das Zurückholen einzelner, versehentlich gelöschter oder durch Softwarefehler beschädigter Dateien – ein Fall, der in der Praxis weit häufiger auftritt als eine komplette Wiederherstellung nach einem Hardware-Ausfall – erfordert bei inkrementellen und differenziellen Sicherungen mehr Aufwand. Unter Umständen muss der Administrator dafür alle Sicherungsbänder einlesen.

Progressive Incremental Backup, manchmal auch Progressive Backup genannt, bietet wie eine Vollsicherung eine schnelle und korrekte Wiederherstellung – auch einzelner Dateien – und gleichzeitig kurze Backup-Zeiten, wie man sie vom Incremental Backup kennt. Ersteres erreicht es durch „Umdrehen“ des Backup-Repository: Die jeweils letzte Sicherung liegt als Full Backup vor, alle älteren als sogenannte Reverse Incrementals.

Außerdem sind alle gesicherten Dateien und Versionen in einer Datenbank gespeichert. Anhand dieses Verzeichnisses kann die Backup-Software feststellen, welche Dateien bereits gesichert sind; sie muss also wie beim Incremental Backup nur die ungesicherten zum Sicherungsserver übertragen. Der übernimmt die Dateien ins Full Backup und verschiebt eventuell vorhandene ältere Versionen in ein neues Reverse Incremental. Ein Flag in der Datenbank zeigt an, ob eine Datei zur letzten Sicherung gehört (aktiv) oder nicht (inaktiv).

Ein weiterer Vorteil des Verfahrens ist, dass man für jedes System nur eine einzige Vollsicherung durchführen muss: zu Beginn, wenn man es in seine Backup-Infrastruktur einbindet. Alle folgenden Sicherungen sind inkrementell (Incremental Forever), solange der Administrator nicht ausdrücklich eine neue Vollsicherung verlangt. Das dürfte allerdings nur selten notwendig sein.

Darüber hinaus erlaubt Progressive Incremental Backup es, nicht mehr benötigte Sicherungen automatisch aus dem Bestand zu entfernen: Der Backup-Server muss lediglich alte Reverse Incrementals löschen und seine Datenbank aktualisieren. Der Administrator hat nichts weiter zu tun als die gewünschte Haltezeit (Retention Time) einzustellen.

Michael Riepe

zierung wegen der Performance-Beeinträchtigung und der Kosten nicht flächendeckend einsetzen kann. Man müsste große zusätzliche Disksysteme hinstellen, die Daten nachts überspielen und anschließend deduplizieren.

Schiebt man die Daten hinterher weiter aufs Tape, muss man sie aus Gründen der Verfügbarkeit wieder duplizieren. Dadurch spare man letztlich nicht viel, meint der Backup-Spezialist von ZAG, denn Deduplizierung sei nur wirksam, wenn alle Daten auf Festplatten liegen, was im Petabyte-Bereich an die Grenzen der verfügbaren Kapazität stoße. Dagegen setze man die Deduplizierungstechnik an kleinen Standorten ein, weil man so den Datenbestand reduzieren könne, bevor man ihn über eine vergleichsweise langsame WAN-Verbindung repliziert.

Hilfreiche Schnappschüsse

Manche Anbieter propagieren Snapshots als Alternative zur Datensicherung. Sie haben jedoch mit Backup nichts zu tun – das ist zumindest die Überzeugung der Backup-Abteilung. Natürlich nutze man auch Snapshots. Ihr Vorteil bestehe vor allem darin, dass man Restores über eine Schnittstelle von Windows direkt an die Fachabteilungen zurückgeben könne: Ältere Versionen lassen sich ohne Umwege reaktivieren. Außerdem nutzt ZAG die Funktion „Snapshop Mirroring“ der Netapp-Filer, um Daten von einem entfernten Standort zu einem anderen zu übertragen.

Im zentralen Rechenzentrum kommen zurzeit noch keine Snapshots zum Einsatz. Man eruiere gerade, ob man sie verwenden kann, um die Systeme beim Backup zu entlasten oder bei logischen Ausfällen ein System schnell wiederherzustellen. Eine weitere Einsatzmöglichkeit wäre, SAP-Systeme zu klonen. Eine Systemherstellung ließe sich über einen solchen Spiegel in ein bis zwei Stunden bewerkstelligen, während sie momentan noch mindestens einen Tag in Anspruch nimmt.

Ziel ist es, Snapshots als Grundlage für die Datensicherung zu nutzen, um das gesicherte Primärsystem nicht zu belasten: Das eigentliche Backup soll in Zukunft von einem Snapshot auf einem Sekundärsystem gezogen werden.

Vom Backup zum Archiv

Bei der Backup-Abteilung von ZAG stellt man in Abrede, dass es genaue Definitionen für Backup und Archivierung gibt. Man habe in der Vergangenheit für sich selbst strenge Regeln erlassen und auf deren Grundlage bestimmt, was Backup ist. Auf der Basis von TSM hat man Aufbewahrungsfristen von einem halben bis zu einem Jahr für Backups festgelegt, je nach Anforderungen der Applikationen. Eine automatische Übernahme in ein Archiv ist strikt untersagt. Ein Projektleiter muss gegebenenfalls das Archivieren bestimmter Datensätze explizit anordnen.

Dafür gibt es zwei Möglichkeiten: ein zertifiziertes revisionssicheres Archivsystem, in dem die Daten unverändert bis zu einem festgelegten Datum liegen, sowie als kostengünstigere Alternative einen sogenannten Langzeitdatenspeicher, der zwar ebenfalls die Integrität der Daten gewährleistet, aber nicht zertifiziert ist. Letzterer speichert die Daten üblicherweise auf Tape, während bei Ersterem unternehmenskritische Daten auf Platte liegen. Außerdem spiegelt ZAG sie auf ein weiteres, 200 Kilometer entfernt stehendes System, um die Verfügbarkeit zu gewährleisten.

Löschoperationen sind in diesem Archivkonzept die Ausnahme. Sie werden nur auf ausdrückliche Anordnung der „Herrn der Daten“ durchgeführt, der Projektverantwortlichen und Leiter der Fachabteilungen. Die IT-Abteilung als Dienstleister hat keine „Besitzrechte“ an den Daten. Auch über deren (Un-)Wichtigkeit entscheiden nicht die IT-Leute. Es wäre kontraproduktiv, wenn die IT-Abteilung etwa „generell unwichtige“ MP3-Dateien automatisch auf günstigere Medien migrieren oder archivieren würde – es gibt auch Audiodaten, die man revisionssicher aufbewahren muss. Aus dem Dateityp beziehungsweise der Namensendung lässt sich nicht ableiten, was man aufbewahren muss und was nicht. Das kann allein der Eigentümer, indem er die Daten bestimmten Klassen zuordnet.

Generell ist ZAG mit der gewählten Lösung zufrieden, zumal sie die Anbindung vieler Betriebssysteme und Anwendungen erlaubt. Allerdings wünscht man sich vom Softwarehersteller IBM eine größere Wertschätzung für das Thema Backup. An die Datensicherung werde immer zuletzt gedacht – eine Haltung, die leider bei den meisten Herstellern zu beobachten sei. Die Entwicklung sei in der Regel zu sehr applikationsgetrieben, das Backup komme zu kurz. Außerdem vermisst der IT-Leiter eine Kooperation zwischen Anwendungs- und Speicheranbietern. Zu oft müsse man als Anwender Druck auf die Hersteller ausüben.

Hartmut Wiehr

In iX extra 01/2009

Networking – Hosting-Provider

Die Branche der Webhoster ist erwachsen geworden. Ihre modernen Rechenzentren übernehmen mehr und mehr IT-Aufgaben im Auftrag von Unternehmenskunden und setzen selbst die Maßstäbe in Sachen Ausfallsicherheit, Energie- und Kosteneffizienz.

Damit ist Hosting heute ein Dienst, der für Großunternehmen genauso infrage kommt wie für den Mittelstand und Privatanwender. Es gilt allerdings, in einem Dschungel von Markennamen und Tarifsystemen den Überblick zu behalten.

Erscheinungstermin:
18. Dezember 2008

DIE WEITEREN IX EXTRAS:

Ausgabe		Thema	Erscheinungstermin
02/09	Embedded Systems	Industrie-PCs und ihre Betriebssysteme	22.01.09
03/09	IT-Security	Identitätsmanagement in heterogenen Netzen	19.02.09
04/09	Storage	Speichertrend: SAN in a box	26.03.09